Sample File:

DSSTox Field Definition File:

EPA Fathead Minnow Acute Toxicity Database (EPAFHM)

(last updated 19 October 03)

Description: Information in this file is intended to provide a minimum level of annotation to the DSSTox SDF files created for the EPA Fathead Minnow Acute Toxicity Database (EPAFHM) obtained from the Source website. For further explanation of Source-specific fields, a user is encouraged to consult the listed references. The Source contact and recommended citation is listed on the DSSTox SDF Download Page (http://www.epa.gov/nheerl/dsstox/)and below. A number of modifications in fields (and allowable contents) were made to the original EPA Fathead Minnow database to improve consistency in notations, eliminate the need for table addendums, and facilitate use of the DSSTox SDF files in relational searching applications. All modifications are fully documented in the ***Comments*** section of the table below. In the course of creating the DSSTox SDF files, it was noted that the Source EPA Fathead Minnow database included a single LC50 test result rather than a geometric mean when multiple (2-10) bioassays were performed on a single chemical. Replicate bioassays were conducted on 98 chemicals in the database. In order to align with the Main Citation - Appendix 2, these replicate experiments are not listed separately in the DSSTox EPAFHM SDF files; rather a single geometric mean of replicate LC50 values for each of the 98 chemicals is reported. As a result of this finding, specific listings of all 137 replicate LC50 experiments for the 98 chemicals have been added to the original EPA Fathead Minnow database offered at the Source Website (see below).

The first section of the table below lists and defines the ***DSSTox Standard Chemical Fields*** used in the EPAFHM SDF files. Any modifications in these fields, deviating either from the original Source data tables or the ***Central List of DSSTox Standard Chemical Fields*** are noted in the ***Comments*** section. Following that section, all Source-specific fields in the EPAFHM SDF files are listed and defined. The ***DSSTox SDF*** column lists SDF files in which the corresponding ***Field Name*** is present. All ***Units*** and ***Descriptions*** are extracted from Source reference materials unless otherwise noted. In some cases, modifications in ***Field Name*** and ***Allowable Values*** from the original data tables were made to facilitate creation and use of the DSSTox SDF files. All differences are noted in the ***Comments*** section. ***Allowable Values*** list allowed field entries occurring in EPAFHM, separated by slashes for exclusive entries (i.e. cannot occur with another entry) and commas or spaces for non-exclusive entries (i.e. can occur with other values). These are defined and explained in the ***Description*** section; italicized note refers to the type of entry (e.g., *Text*); the pound symbol (#) indicates that the ***Allowable Values*** entry is a number. To minimize problems with import and export of SDF files, we avoid the use of punctuation and symbols in ***Allowable Values*** wherever possible. Upper and lower cases in ***Allowable Values*** text entries are used only for emphasis, and not alone to distinguish separate meaning except in the case of SMILES entries, which are case-sensitive.

Source Website: http://www.epa.gov/med/databases/fathead_minnow.htm

Source Contact: Scientific questions pertaining to the EPAFHM database should be directed to Chris Russom, Mid-continent Ecology Division, National Health & Environmental Effects Research Laboratory, US EPA, Duluth, MN; email: russom.chris@epa.gov

Main Citation: Publications reporting use of DSSTox SDF files for the EPA Fathead Minnow database are asked to list the full DSSTox file name(s), including date stamp, and to cite as primary reference the following:

> Russom, C.L., S.P. Bradbury, S.J. Broderius, D.E. Hammermeister, and R.A. Drummond (1997) Predicting modes of action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environmental Toxicology and Chemistry* 16(5): 948-967. *
>
> **pdf of Main Citation can be downloaded from EPAFHM SDF Download Page at the central DSSTox website:* http://www.epa.gov/nheerl/dsstox/

SDF Development Notes:
Each DSSTox SDF file contains a single **Structure** field whose entry corresponds to the **StructureShown**, **CAS**, **SMILES**, **Formula**, and **MolWeight** fields. The main DSSTox SDF file represents the actual tested form of the chemical in the **Structure** field (see ***Description*** below), including complexed molecular entities and salt counter ions in all cases. An additional DSSTox “Defined Organic Parent” SDF file (EPAFHM_DOP) is offered for download for specialized use in

Structure-Activity Relationship (SAR) modeling applications. This file contains no inorganics, organometallics, or mixtures, and all defined organic salts and complexes are stripped of counter-ions and complexed molecular entities and converted to a simplified parent representation in the **Structure** field. The **StructureShown** entry for these compounds is "simplified to parent", with corresponding changes in the **CAS**, **SMILES**, **Formula**, and **MolWeight** field entries. These "simplified to parent" structures are represented in neutralized (protonated) form wherever possible [exceptions include quaternary ammonium and pyridinium ions, which are represented as positively charged (N+) stripped of counter ions, and nitro compounds, which are represented in the charge-separated form, i.e. N+(=O)(O-)]. In the DOP file, both a **CAS_TestedForm** and **SMILES_TestedForm** field are included to allow a user to refer back to the original CAS and SMILES of the tested form of the chemical (i.e., salt or complex). The remaining field contents of the DOP file are identical to that of the main EPAFHM SDF for the subset of "defined organics". Users should be aware that most commercial chemical relational database applications automatically insert one or more structure identifier fields upon import or export of an SDF file (e.g., FW or Mol_ID), fields that may augment or duplicate one or more of the DSSTox Standard Chemical Fields. Also, since the proper ordering of fields upon SDF import into most applications requires a non-blank entry in each field of the first database record, the word "blank" is entered in each empty text field in Record 1 for this purpose; this word can be deleted from Record 1 fields after SDF import is complete.

As an MS Word document, the following table is best viewed onscreen using either Normal or Web Layout View in Landscape page orientation. Page breaks have been inserted in both the MS Word and PDF versions of this table for optimal page layout view and printing.

Field Name	*DSSTox SDF*	*Units*	*Allowable Values*	*Description*	*Comments*
DSSTox Standard Chemical Fields					
Structure	All		*Molecule*	Two-dimensional graphical representation of molecular structure. Form of structure is identified in the **StructureShown** field and always corresponds to the fields: **CAS**, **SMILES**, **MolWeight**, **Formula**. Structure entry is *blank* when **SubstanceType** entry is "mixture or unknown".	Three tested substances (Demeton, t-butyl styrene and chloromethyl styrene) are isomer mixtures that were represented by a single isomer SMILES code in the original EPA Fathead Minnow database. Here they are classified as mixtures and hence excluded from the "DOP" file.
Structure Shown *(no spaces)*	All		tested form/ simplified to parent/	Identifies form of graphical 2D structure displayed in the **Structure** field. Field entry is "tested form" for all records in main SDF. For DOP file, entry is "simplified to parent" for "defined organic" salts and complexes.	
Formula	All		*Text*	Empirical formula of displayed **Structure**.	
MolWeight	All	amu	#	Molecular weight of displayed **Structure**.	Same content as in field "MW" provided in original EPA Fathead Minnow database.
CAS	All		NOCAS/ ######-##-#/	Chemical Abstracts Service (CAS) Registry number of displayed **Structure**, formatted **000000-00-0**, corresponds to **StructureShown** and **SMILES**. "NOCAS" indicates CAS number was unavailable from original Source data table or was not found.	In a few cases, where tested form CAS did not correspond exactly to NAME in original database and the latter was more specific, DSSTox CAS for tested form was changed to correspond to NAME.
SMILES	All		*Text*	SMILES molecular text code of displayed **Structure**, corresponding to **StructureShown** and **CAS**.	
DSSTox_ID	All		#	Sequential ID number assigned to each record in database, values range from 1 to n= total #records. When accompanied by full DSSTox filename, provides unique record identifier.	

DSSTox_ FileName	All		*Text*	Full DSSTox SDF standard file name without .sdf extension. Field entry will be updated whenever new version or revision of SDF database is generated.	e.g., EPAFHM_v1a_617_15Oct03 --or-- EPAFHM_DOP_v1a_610_15Oct03
ChemName	All		*Text*	Common or trade name of chemical listed in original Source data table, corresponds to the original tested form of the chemical or substance, if known.	Same content as in field "Name" provided in original EPA Fathead Minnow database.
Substance Type *(no spaces)*	All		defined organic/ inorganic/ organometallic/ mixture or unknown/	Nature of chemical or substance: "defined organic" = defined chemical structure containing carbon but not organometallic, i.e. with no metal or metalloid other than simple salt alkali (I) or alkali earth (II) metals; "inorganic" = defined chemical structure containing no carbon; "organometallic" = defined chemical structure containing carbon and any metal or metalloid other than alkali (I) or alkaline earth (II) metals (such as Na, K, Mg, Ca) that occur in simple salts ; organometallics always labeled "complex" in **TestedForm** field; "mixture or unknown" could represent mixture, ill-defined, or unknown substance type, has no entry in **Structure**, **StructureShown**, **TestedForm**, and **SMILES** fields.	Information inferred from "Name" and "CAS" entries listed in original EPA Fathead Minnow database.
TestedForm	All files containing tested substances		parent/ salt, complex/	Tested form of chemical inferred from **ChemName**, **CAS** and **Structure**, with DSSTox operational definitions as follows: "parent" = single chemical entity, without counter ions or complexed chemical entities; "salt" = simple ionic salts of alkali (I) or alkaline earth (II) metals (such as Na, K, Mg, Ca) or halides (e.g., Cl, Br); "complex" = any compound with associated acid, base, or hydrate, or any organometallic.	Tested form of chemical inferred from "Name" and "CAS" entries listed in original EPA Fathead Minnow database.
AddToParent	All files containing salts or complexes		*Text*	For **SubstanceType=**"defined organic" and **TestedForm=**"salt" or "complex", entry specifies salt counter-ions or complexed entities (e.g., Na, K, HCl, Cl, H2O, Ca , H2SO4, acetate, etc.) that are removed when **StructureShown**="simplified to parent" in DOP file; ; "bis" signifies parent structure occurs twice in complex.	
ChemNote	All		*Text,* quaternary ammonium, stereochem, replicate 2D, replicate parent, tautomeric form, *etc.*	Note related to nature of chemical in exceptional cases, e.g., if uncertainty exists in **ChemName** or **CAS**, if parent structure is "quaternary ammonium" ion, if mixture characteristics are known, if "stereochem" information is known (e.g., *cis, trans, E,Z*, R, S), etc. Note indicates replicate parent or 2D structures in database. Note also indicates if structure is one of two tautomeric forms.	For tested substances Demeton, t-butyl styrene and chloromethyl styrene, mixture details with SMILES for both isomers provided. Replicate parent structures for phenyltrimethylammonium-iodide and –methosulfate, and replicate 2D structures for cis- and trans-3-hexen-1-ol.

CAS_ TestedForm	Files containing simplified structures, e.g., DOP		NOCAS/ ######-##-#/	CAS of actual tested form of chemical, formatted **000000-00-0**, differs from **CAS** field entry only when a simplified form of chemical is represented in the **Structure** field. "NOCAS" indicates CAS number was not found.	Field occurs only in DOP file and entry differs from **CAS** only for salts and complexes when **StructureShown**="simplified to parent".
SMILES_ TestedForm	Files containing simplified structures, e.g., DOP		*Text*	SMILES molecular text code of actual tested form of chemical, differs from **SMILES** field entry only when a simplified form of chemical is represented in the **Structure** field. Corresponds to **CAS_TestedForm**.	Field occurs only in DOP file and entry differs from **SMILES** only for salts and complexes when **StructureShown**="simplified to parent".
EPAFHM Source-Specific Fields					
ChemClass FHM	EPAFHM	*None*	Alkanes/ Alkenes/ Saturated Hydrocarbons/ Unsaturated Hydrocarbons/ Basic Ethers/ Diphenyl Ethers/ Cyclic Ethers/ Basic Alcohols/ Alkene Alcohols/ Alkyne Alcohols/ Diols/ Aldehydes/ Basic Ketones/ beta Diketones/ Cyclic Ketones/ Carboxylic Acids/ Basic Esters/ Phthalates/ Amides/ Acrylates/ Nitriles/ Primary aliphatic amines/ Secondary aliphatic amines/ Tertiary aliphatic amines/ Primary aromatic amines/ Secondary aromatic amines/ Tertiary aromatic amines/ Azine compounds/ Sulfides/ Disulfides/ Sulfo compounds/ Benzenes/ Chlorinated Benzenes/ Phenols/ Chlorinated Phenols/ Piperazines/ …..	Standard organic chemical class designations of the sort used in traditional QSAR studies. These class designations are only provided for information purposes in EPAFHM and were not used in the construction of MOA classes or derivation of QSARs for this study.	To facilitate use of EPA Fathead Minnow database in a relational format, we have replaced the original field "CODE" with new field "ChemClass FHM", which contains the organic chemical class referenced in the "CODE" field and listed in a Table 2 addendum in the original EPA Fathead Minnow database. DSSTox field entries exactly correspond to original EPA Fathead Minnow database CODE Table 2 entries with the exception that commas are deleted from Primary, Secondary, Tertiary, aliphatic and aromatic amines class names, and hyphens are deleted from 5-membered ring aliphatics and aromatics, beta-Diketones, hetero-atom compounds, and DEAS-complex structures class names. DEAS stands for "Drug Enforcement Agency Structures", a broad functional categorization.

CLOGP	EPAFHM	*None*	#	Logarithm of the octanol:water partition coefficient (LogP) computed using the semiempirical fragment-based method applied in the CLOGP software [2], unless M appears in MLOGP field in which case the measured LogP value is provided in this field from the STARLIST database of CLOGP.	
….					
LC50	EPAFHM	mg/l	*blank/* #	96 h LC50 (concentration producing lethality in 50% of test animals after 96 hours exposure) in mg/l. Calculated using Spearman-Karber method [3]. Geometric mean of LC50s presented if more than one bioassay conducted for the chemical. If no mortality, or less than 50% mortality observed at 96 h, LC50 field left *blank*.	If insufficient mortality was observed, LC50 was assigned a value of –99 in the original EPA Fathead Minnow database. To avoid problems when log(LC50) is used in QSAR, we convert these to *blank* entry in LC50 field. Replicate experiments used to compute geometric mean LC50 values for 98 chemicals are listed in revised EPA Fathead Minnow database located at Source website. Geometric mean LC50 value for pentachlorophenol is reported in DSSTox SDF.
LC50NOTE	EPAFHM	*None*	*Text*	Comments regarding the LC50 result, pertaining to exceptional situations, e.g., where: 50% mortality could not be achieved at saturation concentrations, non-monitonical pattern of death was observed (i.e., more deaths at lower concentrations than at higher concentrations), or exceptions were made in terms of pH or mixtures. If more than one replicate bioassay, number of experiments contributing to calculation of the geometric mean LC50 is specified.	Replaces REMARKS field in original EPA Fathead Minnow database; eliminate use of abbreviations for mortality and saturation. When more than one bioassay was performed on chemical, note refers to single test result. A note was added to the DSSTox SDF specifying the number of experiments (2-10) contributing to calculation of the geometric mean LC50.
File truncated					

Additional EPAFHM references:

1. Anderson, E., G.D. Veith, and D. Weininger (1987) SMILES: A line notation and computerized interpreter for chemical structure. EPA/600/M-87-021. Technical Report. U.S. Environmental Protection Agency, Environmental Research Laboratory, Duluth, MN, USA.

2. CLOGP™ program version 3.4 and STARLIST database, respectively, within the UDRIVE system version 3.53, 1988, from Pomona College Medicinal Chemistry Project, Claremont, CA.

3. Hamilton, M.A., R.C. Russo, and R.V. Thurston (1977) Trimmed Spearman-Karber method for estimating median lethal concentrations in toxicity bioassays. *Environ. Sci. & Technol.* 11: 714-719. Correction 12: 417.

4. Broderius, S., M. Kahl, and M. Hoglund (1995) Use of Joint Toxic Response to define the primary mode of toxic action for diverse industrial organic chemicals. *Environ Toxicol Chem* 14: 1591-1605.

….